Cell, Vol. S116, S75–S76, January 23, 2004 Copyright ©2004 by Cell Press

Actively Seeking Activating Sequences Commentary

Jun Ma*
Division of Developmental Biology
Cincinnati Children's Hospital Research Foundation
University of Cincinnati College of Medicine
3333 Burnet Avenue
Cincinnati, Ohio 45229

In the spring of 1986, I was recovering from yet another disappointment. Mark Ptashne, my Ph.D. supervisor, had been actively promoting the looping model for how regulatory proteins act at a distance on DNA. Inspired by the λ repressor work of Ann Hochschild (1986), another student in the lab, I wanted to determine whether the yeast activator Gal4 had to sit on the same side of the DNA helix as a nearby TATA box to activate transcription, as predicted by the looping model. But my experiments showed that Gal4 worked equally well regardless of whether its binding site and the TATA box were on the same or opposite side of the DNA helix. Reflecting on this result, I thought it might be due to the modular—perhaps highly flexible—nature of this large protein of 881 aa. Recent experiments by Roger Brent (1985), a postdoctoral fellow in the lab, and two students, Liam Keegan and Grace Gill (1986), had just shown that while the N-terminal 74 aa of Gal4 were responsible for bringing the protein to DNA, the remaining 90% of the protein provided an activation function. I began to wonder if I could generate smaller Gal4 derivatives to determine whether they might exhibit the anticipated helical periodicity.

After a few pilot tests, I soon initiated a systematic deletion analysis of Gal4, progressively removing protein sequences both internally and from the C terminus. The urgency of this analysis was also fueled by a rumor that Keith Yamamoto's lab was able to generate large truncations of the glucocorticoid receptor without destroying its activation function. My experiments (1987a) identified two short segments of Gal4 (49 and 114 aa), each sufficient to activate transcription when linked to its DNA binding domain. Derivatives of Gal4 lacking as much as 80% of their protein sequence can activate transcription. Later, Doug Ruden (1988), another student in the lab, and I looked for helical periodicity using small derivatives of Gal4, but we never found any.

Before the completion of the Gal4 deletion analysis, I read a report by Ian Hope and Kevin Struhl (1986), who had delineated a 19 aa activating region for another yeast activator Gcn4. Interestingly, the two Gal4 activating sequences and that of Gcn4 shared no obvious homology among themselves except being highly acidic. As noted by Hope and Struhl, the Gcn4 activating region also failed to show sequence homology to other yeast activators available at the time. An idea began to emerge that many different sequences—perhaps sharing some loosely defined common features such as an excess of acidic residues—could activate transcription in yeast, presumably by interacting with some proteins in the transcription machinery.

To test this idea, I constructed a special library by fusing random bits of *E. coli* genomic DNA to the coding sequence for the Gal4 DNA binding domain (1987b). When I introduced the library into a *gal4*$^-$ yeast strain carrying a Gal4-responsive *lacZ* reporter gene, I found that a surprisingly high percentage (~1%) of the colonies turned blue on X-gal plates. After confirming that the blue (*lacZ*$^+$) phenotype was linked to library plasmids, anxiety-turned-excitement became evident on a transparency I showed at a lab meeting: "No Hoax!", it declared. I strategically chose 15 shortest *E. coli* fragments for Maxam-Gilbert DNA sequencing to localize, effectively, the encoded activating sequences, referred to as B sequences (for blue, not blob). The B sequences vary in length, and the shortest ones have fewer than 20 aa. While most of the B sequences differ from one another and none has obvious homology to those of Gcn4 and Gal4, they are all characteristically acidic. In a concurrent mutagenesis analysis of one of the activating regions of Gal4, Grace (1987) further showed that acidity is important for activation. (Earlier, at around the time of my deletion analysis, Grace had performed another genetic study to isolate full-length Gal4 mutants specifically defective in activation, but only frameshift mutations appeared.)

One of the B sequences, B42, also activated transcription when fused to the bacterial LexA DNA binding domain. This finding completely demystified eukaryotic activators, reinforcing the notion put forth earlier by Roger, Liam, and Grace that an activator requires two, and only two, functions: one for directing it to DNA and the other for activating transcription. Evidently, neither function has to be conferred by proteins of eukaryotic origin. Since activating sequences are so "simple" and so easy to find, it became virtually impossible to imagine how activators might work by exerting some enzymatic activities themselves, rather than by simply touching other proteins.

Experiments with B42 further revealed that an activating sequence does not have to be directly linked to a DNA binding domain. At the San Francisco yeast meeting in June 1987, I heard a talk by Yasuhisa Nogi that a region of Gal80 could tolerate insertions and deletions. Gal80 is a Gal4-specific repressor believed to inhibit Gal4's activity by masking its activating surfaces. Nogi's report immediately inspired me to ask whether insertion of an activating sequence to Gal80 could convert it to an activator. My experiments showed (1988) that a Gal80-B42 hybrid protein activated transcription if, and only if, yeast cells also had a Gal4 derivative capable of interacting with both Gal80 and DNA. On a ski lift during a lab ski trip that winter, Mark and I discussed my results and enjoyed the concept of experimentally bringing an activating sequence to DNA through protein-protein interactions. This concept was later extended by Stanley Fields (1989), who proposed the yeast two-hybrid system.

Subsequent studies have since revealed that acidic

*Correspondence: jun.ma@cchmc.org

activating sequences and their targets interact in much the same way as bees and flowers in a bouquet. First, an acidic activating sequence tends to have the ability to interact with multiple target proteins in the transcriptional apparatus. In retrospect, this may explain why Gal4, and even its small derivatives, failed to exhibit the anticipated helical periodicity: it is likely that target surfaces are available from multiple sides of the DNA helix. Second, different acidic activating sequences tend to interact with the same (or overlapping sets of) target proteins. Such an interaction mode is reflective of what the acidic activating regions are: short sequences sharing loosely defined common features but nevertheless relying on some structural characteristics for function. This mode is distinct from its "dedicated partner" alternative, in which activating sequences and targets could only interact in a pair-wise manner. It offers a vast interaction repertoire between activators and targets. This enables the transcriptional apparatus to integrate activation inputs, in all necessary combinations, in response to environmental and developmental cues. The invention of nonacidic types of activating sequences (and their targets) in higher eukaryotes further expands the activator-target interaction repertoire to meet greater demands.

Acknowledgments

I thank R. Brent, I. Cartwright, L. Keegan, Z. Liu, M. Ptashne, and M. Xu for comments. I apologize that the page limit did not permit citation of all relevant work. This work was supported in part by the AHA and the NSF.

References

Brent, R., and Ptashne, M. (1985). A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. Cell *43*, 729–736.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245–246.

Gill, G., and Ptashne, M. (1987). Mutants of GAL4 protein altered in an activation function. Cell *51*, 121–126.

Hochschild, A., and Ptashne, M. (1986). Cooperative binding of λ repressors to sites separated by integral turns of the DNA helix. Cell *44*, 681–687.

Hope, I.A., and Struhl, K. (1986). Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. Cell *46*, 885–894.

Keegan, L., Gill, G., and Ptashne, M. (1986). Separation of DNA binding from the transcriptional-activating function of a eukaryotic regulatory protein. Science *231*, 699–704.

Ma, J., and Ptashne, M. (1987a). Deletion analysis of GAL4 defines two transcriptional activating segments. Cell *48*, 847–853.

Ma, J., and Ptashne, M. (1987b). A new class of yeast transcriptional activators. Cell *51*, 113–119.

Ma, J., and Ptashne, M. (1988). Converting a eukaryotic transcriptional inhibitor into an activator. Cell *55*, 443–446.

Ruden, D., Ma, J., and Ptashne, M. (1988). No strict alignment is required between a transcriptional activator binding site and the "TATA box" of a yeast gene. Proc. Natl. Acad. Sci. USA *85*, 4262–4266.

Cell, Vol. 43, 729–736, December, 1985 (Part 2), Copyright ©1985 by Cell Press

A Eukaryotic Transcriptional Activator Bearing the DNA Specificity of a Prokaryotic Repressor

Roger Brent* and Mark Ptashne
Department of Biochemistry and Molecular Biology
Harvard University
7 Divinity Avenue
Cambridge, Massachusetts 02138

Summary

We describe a new protein that binds to DNA and activates gene transcription in yeast. This protein, LexA-GAL4, is a hybrid of LexA, an Escherichia coli repressor protein, and GAL4, a Saccharomyces cerevislae transcriptional activator. The hybrid protein, synthesized in yeast, activates transcription of a gene if and only if a *lexA* operator is present near the transcription start site. Thus, the DNA binding function of GAL4 can be replaced with that of a prokaryotic repressor without loss of the transcriptional activation function. These results suggest that DNA-bound LexA-GAL4 and DNA-bound GAL4 activate transcription by contracting other proteins.

Introduction

In Saccharomyces cerevisiae, the protein GAL4 turns on transcription of the *GAL1* gene when bound to an upstream region called UAS_G. This region contains four 17 bp sites of related sequence, a near-consensus of which (the "17-mer") mediates GAL4 activity in vivo and binds GAL4 in vitro (Giniger et al., 1985; Keegan, personal communication). Both UAS_G and a single 17-mer function when placed at several positions within a region between 40 and 600 nucleotides from the *GAL1* transcription start site, or when placed upstream of a different gene, *CYC1* (Guarente et al., 1982b; West et al., 1984; Giniger et al., 1985). GAL4 is active in wild-type strains only when cells are grown on medium containing galactose, because, it is thought, growth on this medium leads to dissociation of GAL4 from an inhibitory protein, GAL80 (Oshima, 1982). GAL4 activity is reduced when cells are grown on medium that contains glucose and galactose (Oshima, 1982; Yocum et al., 1984), at least partly because GAL4 binds UAS_G inefficiently under these conditions (Giniger et al., 1985).

Upstream activation sites (UASs) have been found upstream of all RNA polymerase II-dependent yeast genes the regulatory regions of which have been carefully studied. For example, upstream of *CYC1* are two sites called UAS_{C1} and UAS_{C2}. Cellular gene products, probably encoded by the *HAP1, HAP2,* and *HAP3* genes (Guarente et al., 1984), presumably interact with these sites. If UAS_G is inserted upstream of *CYC1* in place of UAS_{C1} and UAS_{C2}, *CYC1* transcription becomes dependent on GAL4 and is regulated like *GAL1* transcription. Although the properties of UASs are similar in other respects to those of the enhancer sequences found in higher eukaryotes, UAS_G and the two UAS_Cs have been reported to be inactive when positioned downstream of the transcription start point of a gene (Struhl, 1984; Guarente and Hoar, 1984).

The current investigation was prompted by a consideration of two mechanisms by which GAL4 might turn on transcription. According to the first, GAL4 would bind to DNA in some way that would stabilize an unusual DNA structure (eg., left-handed DNA), and the perturbed structure would then somehow be transmitted down the helix, where it would help proteins bind near the transcription start. According to the second idea, GAL4 would contact DNA without greatly perturbing the structure of the DNA around the binding site, and activation of transcription would occur when GAL4 touches other proteins. In Escherichia coli, lambda repressor acts as a positive regulator (of its own gene) by the second mechanism; repressor binds to a site adjacent to the RNA polymerase binding site and touches RNA polymerase. One line of evidence that led to this picture was the isolation of lambda repressor mutants called *pc* (for Positive Control) that bind DNA but fail to activate transcription (Guarente et al., 1982a). The amino acids changed in *pc* mutants are clustered in a region on the surface of the lambda repressor molecule (Hochschild et al., 1983) that is thought on the basis of other experiments to be that portion of the molecule that touches RNA polymerase.

Consideration of the lambda experiments led us to try to separate the ability of GAL4 to bind DNA from its ability to stimulate transcription. However, instead of seeking to preserve GAL4's DNA binding while eliminating its ability to activate transcription, we sought to confer a new DNA binding specificity on GAL4 while preserving its ability to stimulate transcription. To this end, we constructed a new protein called LexA-GAL4, the DNA binding specificity of which came from an E. coli repressor called LexA.

In E. coli, LexA represses many genes. Like the repressors of lambda-like phages, LexA probably binds as a dimer to its operators (R. Brent, Ph.D. thesis, Harvard University, Cambridge, Massachusetts, 1982). Moreover, the LexA monomer seems to have an overall organization similar to that of the phage repressors; an amino terminal domain that binds operator DNA and contains weak dimerization contacts, a carboxy-terminal domain that contains stronger dimer contacts, and a flexible hinge region that connects the two (Brent and Ptashne, 1981; R. Brent, Ph.D. thesis 1982; Little and Hill, 1985; Shnarr et al., 1985). The first 87 amino acids of LexA contain the information necessary for specific binding to the LexA operator (Brent, unpublished) and 16 amino acids of the putative hinge region (Little and Hill, 1985). If the cellular DNA is damaged, RecA protein and amino acids within the C-terminus of LexA catalyze cleavage of LexA within the hinge region.

* Present address: Department of Molecular Biology, Massachusetts General Hospital, 50 Blossom Street, Boston, Massachusetts 02114, and Department of Genetics, Harvard University Medical School, Boston, Massachusetts 02115.

Cell, Vol. 51, 113–119, October 9, 1987, Copyright ©1987 by Cell Press

A New Class of Yeast Transcriptional Activators

Jun Ma and Mark Ptashne
Department of Biochemistry and Molecular Biology
Harvard University
Cambridge, Massachusetts 02138

Summary

We describe yeast transcriptional activators encoded by E. coli genomic DNA fragments fused to the coding sequence of the DNA-binding portion of GAL4. All of the new activating sequences that we have analyzed, like those of GAL4 and GCN4, are acidic; most of these sequences show no obvious sequence homology when compared with the identified activating regions of GAL4 and GCN4 or among themselves. We also describe a fusion protein that contains no yeast protein sequence but activates transcription in yeast.

Introduction

In the yeast Saccharomyces cerevisiae the genes required for galactose catabolism (*GAL1*, *GAL7*, and *GAL10*) are expressed at high levels when cells are grown in medium containing galactose (Hopper et al., 1978; Hopper and Rowe, 1978; St John and Davis, 1981). GAL4, a protein of 881 amino acids (Laughon and Gesteland, 1984; Laughon et al., 1984), is required for the expression of these genes (Douglas and Hawthorne, 1964; Oshima, 1982). The protein binds specifically to an upstream activation sequence (called the UAS_G), located midway between the divergently transcribed *GAL1* and *GAL10* genes, and activates transcription (Guarente et al., 1982; Yocum et al., 1984; West et al., 1984; Johnston and Davis, 1984; Bram and Kornberg, 1985; Giniger et al., 1985; Lohr and Hopper, 1985; Selleck and Majors, 1987a, 1987b). The activity of GAL4 is inhibited by the negative regulator GAL80 when cells are grown in medium lacking galactose (Douglas and Hawthorne, 1964, 1966; Oshima, 1982; Torchia et al., 1984; Nogi et al., 1984; Nogi and Fukasawa, 1984). Glucose represses the expression of the *GAL* genes whether or not galactose or GAL80 is present, a phenomenon called glucose repression (Adams, 1972; Matsumoto et al., 1981, 1983).

Two lines of evidence show that GAL4 binds DNA using its amino terminus and activates transcription using other parts of the protein. First, amino terminal fragments of GAL4 (residues 1–98 or 1–147) bind to DNA both in vivo and in vitro but fail to activate transcription (Keegan et al., 1986; Ma and Ptashne, 1987a; J. Ma, H. Kakidani, and M. Ptashne, unpublished data). Second, a fusion protein bearing the DNA-binding domain of the bacterial repressor LexA in place of that of GAL4 activates transcription from a *lexA* operator positioned upstream of a yeast gene (Brent and Ptashne, 1985). More recent experiments show that either of two short regions of the protein (activating region I, residues 148–196 or activating region II, residues 768–881) activates transcription when fused to the DNA-binding portion of the molecule (residues 1–147; Ma and Ptashne, 1987a). The carboxy-terminal 30 amino acids of GAL4, a part of activating region II, are also involved in inhibition by GAL80 (Ma and Ptashne, 1987b; Johnston et al., 1987).

The two activating regions of GAL4, each highly acidic, bear no obvious sequence homology to one another (Ma and Ptashne, 1987a), or to the activating region (also acidic) of another yeast activator, GCN4 (Hope and Struhl, 1986; Struhl, 1987). Protein sequences that bear little sequence homology but nevertheless manifest identical biological functions have been observed in other cases. For example, signal sequences that direct the secretion of proteins bear little, if any, sequence homology (reviewed by Briggs and Gierasch, 1986; also see von Heijne, 1985); the same is true for signal sequences of mitochondrial proteins (von Heijne, 1986; Schatz, 1987). Recently, Kaiser et al. (1987) and Baker and Schatz (1987) replaced the signal sequences of secreted and of mitochondrial proteins, respectively, with random peptide sequences and found that a large fraction of these random sequences can function as signal sequences.

In this paper we describe a new class of yeast activators; these activators were encoded by genes bearing E. coli genomic DNA fragments fused to the coding sequence of the DNA-binding portion of GAL4. We detected these activating sequences at a high frequency, and most of the new activating sequences show no obvious sequence homology when compared with one another or with the activating regions of GAL4 and GCN4. All of the new activating sequences that we have analyzed are acidic. We also show that a fusion protein bearing one of our new activating sequences attached to the DNA-binding domain of LexA—a protein therefore containing no yeast protein sequence—activates transcription in yeast.

Results

Experimental Design

We fused E. coli genomic DNA fragments (generated with the restriction enzyme Sau3A) to the coding sequence of the DNA-binding portion of GAL4 (amino acids 1–147). Expression of these genes, carried on plasmids (Figure 1), was directed by the *ADH1* promoter. We introduced these plasmids into a yeast "tester" strain lacking functional GAL4 but bearing an integrated *GAL1-lacZ* fusion gene with the UAS_G at its normal location. The product of the *GAL1-lacZ* fusion gene (β-galactosidase) conferred a distinctive blue color to the yeast colonies on the appropriate indicator plates (X-gal plates); in the absence of a transcriptional activator that recognizes UAS_G the colonies were white. We detected 154 blue colonies among about 15,000 transformants. We isolated plasmids from 35 individual transformants expressing various levels of β-galactosidase. We retransformed 26 plasmids into the tester strain and found that the β-galactosidase-positive pheno-